UNIVERSIDAD AUTONOMA DE MADRID
ESCUELA POLITECNICA SUPERIOR SUPERIOR

Reconocimiento de escenas exteriores mediante redes neuronales profundas entrenadas con la base de datos Places

Autor: Santiago Vicente Moñivar
Tutor: Miguel Ángel García García

TRABAJO DE FIN DE GRADO

Escuela Politecnica Superior
Universidad Autonoma de Madrid
September 2019

Reconocimiento de escenas exteriores mediante redes neuronales profundas entrenadas con la base de datos Places

Santiago Vicente Moñivar

Tutor: Miguel Ángel García García

Video Processing and Understanding Lab
Escuela Politecnica Superior
Universidad Autonoma de Madrid
September 2019

Trabajo parcialmente financiado por el Ministerio de Economía y Competitividad del Gobierno de España bajo el proyecto TEC2017-88169-R (MobiNetVideo) (2018-2020)

Resumen

El siguiente Trabajo de Fin de Grado se ha basado en dos pilares fundamentales. Uno de ellos es la creación de una base de datos de escenas exteriores, la cual se utilizará para testear con la red neuronal de PLACES365, y otro la modificación de una red para realizar un entrenamiento y posterior testeo con el ya mencionado dataset.

En primer lugar, se ha generado una base de datos con diez clases diferentes y formada por imágenes exteriores de edificios repartidos a lo largo del campus de la Universidad Autónoma de Madrid. Después se ha realizado un estudio de la red neuronal convolucional PLACES365, en concreto ResNet50, implementada y entrenada por el MIT, aplicando nuestra base de datos.

Por otro lado, se ha hecho "transfer learning", modificado la red ya citada, añadiéndole tres nuevas capas al final, con el objetivo de adaptarla a la nueva base de datos siendo capaz de clasificar entre las diez clases creadas para las distintas facultades.

Finalmente, se ha hecho un estudio comparativo entre los distintos resultados obtenidos con el testeo, analizando el motivo potencial de los fallos que se han generado.

Palabras Clave

Red neuronal convolucional, base de datos, entrenamiento, testeo, PLACES365.

Abstract

The following Final Degree Project has been based on two fundamental pillars. One of them is the creation of a database of exterior scenes, which will be used to test the PLACES365 neural network , and the other the modification of a network in order to perform a training and a subsequent testing with the aforementioned dataset.

First, a database with ten different classes has been generated and and constituted by exterior images of buildings distributed over the campus of the Autonomous University of Madrid. Afterwards, a study of the PLACES365 convolutional neural network has been carried out, specifically ResNet50, implemented and trained by the MIT, by applying our database.

On the other hand, transfer learning has been done, modifying the aforementioned network, by adding three new layers at the end, with the aim of adapting it to the new database, being able to classify among the ten classes created for the different faculties.

Finally, a comparative study has been made between the different results obtained after testing, analyzing the potential reason for the failures that have been generated.

Keywords

Convolutional Neural Networks, dataset, train, test,PLACES365

Agradecimientos

Agradecerle primero a mi tutor y consejero Miguel Ángel que me ha dado el apoyo y las herramientas para enfrentarme a este nuevo reto, el que ha sido y será siempre mi primer maestro con el que he abierto vía en mi futuro con la inteligencia artificial.

También me gustaría agradecérselo a mis amigos de la universidad, con los que hace ya bastante tiempo comencé mis andaduras por este camino duro y pedregoso que es la ingeniería, con los cuales espero seguir pudiendo compartir nuevos retos y aventuras en el futuro.

A mis amigos de toda la vida, con los que he crecido y sigo creciendo, que han sabido apoyarme en todos los momentos.

A Ust, que si tuviese que describir todo lo que me ha ayudado en mi vida casi ocuparía mas que este tfg.

Y por último, a mi familia, con la que me he enfrentado a viento y marea, que siempre ha estado ahí para darme el cariño y amor pese a la gran cantidad de veces que les pueda haber fallado.

Santiago Vicente Monivar

Septiembre 2019.

Índice general

Índice de figuras

Capítulo 1

Introducción

1.1. Motivación

El trabajo con redes neuronales convolucionales (CNN) es una parte fundamental de la inteligencia artificial que permite seguir avanzando en el proceso de automatización de las distintas actividades realizadas por las máquinas. Esto nos impulsa a tratar de conocer en mayor medida este campo, todavía de corta de edad, pero con un futuro muy prometedor.

La primera motivación de esta trabajo de fin de grado es estudiar el comportamiento de una red neuronal ya existente, modificándola para que que no tome imágenes de un archivo predefinido, sino que se adapte a un dataset propio. Con esto a parte de lograr ver la precisíon de la CNN con imágenes de un sitio específico como la Universidad Autónoma de Madrid, ajenos a los propuestos por la propia red, también se observará cómo se comporta ante escenas similares entre si, viendo de esta manera, si sería necesario o no realizar otro posterior entrenamiento con un nuevo dataset.

Otra motivación es aprender a implementar, modificar y adaptar una red neuronal convolucional para un problema determinado. Es por ello que la segunda parte del trabajo se centrará en modificar una red mediante transfer learning, siendo necesario aprender su funcionamiento y composición para poder modificarla de manera que seleccione una serie de categorías sobre las que se realizaran posteriores estudios.

Todo este trabajo también tiene como motivación aprender a trabajar con un proyecto de mayor escala a todos los realizados anteriormente en el grado. Esto implica aprender a formarnos en elementos fundamentales como: aprendizaje de forma autodidacta, al enfrentarnos a problemas con un nuevo lenguaje de programación; a cumplir un calendario y manejar los tiempos, aprendiendo así como se trabaja en las empresas, donde se realizan proyectos en periodos específicos de tiempo; y sobre

todo a la perseverancia, que te impulsa a seguir trabajando en momentos de absoluta desesperación.

1.2. Objetivos

Este trabajo de fin de grado tiene varios objetivos principales, todos ellos centrados en las redes neuronales y su comportamiento con distintas bases de datos. Por ello se podría resumir en que el objetivo fundamentas es el aprendizaje y comprensión del funcionamiento de una red neuronal, pudiendo así someterla a distintas pruebas.

El primer objetivo se centra en comprender una red ya creada como Places365, y modificarla de tal manera que realice el testeo con una base de datos nueva que habremos generado. Con esto se busca empezar a conocer el funcionamiento de las redes neuronales, además de realizar un estudio para ver la eficacia real ante una serie de escenas alternativas a las que plantea originariamente el MIT con esta red.

Otro de los objetivos fundamentales es modificar una red convolucional. Para esta modificación se parte de la red ya preentrenada y se busca añadir una serie de capas nuevas, que nos permita cambiar la categorización que trae prediseñada la red, para que en su lugar se vea obligada a entrenar con respecto a diez categorías específicas de escenas exteriores de la universidad Autónoma.

Finalmente se pondrá a prueba esta nueva red, entrenándola con nuestra base de datos y se realizara un estudio para ver el nivel de acierto que tiene con nuevas imágenes de testeo. Con esto se busca analizar el comportamiento real de una red neuronal, entrenada con una base de imágenes mucho menor al que está acostumbrada y ver el nivel de efectividad que nos aporta.

1.3. Estructura de la memoria

- Capítulo 1: Introducción: motivación y objetivos.
- Capítulo 2: Estado del arte.
- Capítulo 3: Diseño.
- Capítulo 4: Desarrollo.
- Capítulo 5: Resultados.
- Capítulo 6: Conclusiones y trabajo futuro.
- Bibliografía.

Capítulo 2

Estado del arte

2.1. Introducción

A lo largo de este capítulo vamos a exponer las tecnologías que hemos utilizado para la realización de este trabajo. Por ello hablaremos de elementos fundamentales como las propias redes neuronales, entorno a las cuales se enfoca todo nuestro trabajo. Por otro lado, también hablaremos de elementos sin los cuales la realización del trabajo habría sido casi imposible, o por lo menos mucho más compleja, como son CUDA y la GPU. Y finalmente hablaremos de los distintos entornos de desarrollo con los que se podría realizar este trabajo, centrándonos en PyTorch, que es el de mayor interés para este trabajo, ya que esta librería es la que nos ha permitido trabajar con la red neuronal.

2.2. GPU

La unidad de procesamiento grafico o GPU (graphics processing unit)[1] sirve como coprocesador cuya función es realizar operaciones de coma flotante y realizar el procesamiento de gráficos, permitiendo así aportar mucha más velocidad y ayudando con la gran cantidad de trabajo que tendría el procesador central al ejecutar las redes neuronales utilizadas en este trabajo de fin de grado.

Se puede encontrar las GPU en algunas tarjetas gráficas de algunos ordenadores muy potentes, o como se ha realizado en este proyecto, a través de algunos entornos virtuales. La GPU incrementa las operaciones primitivas optimizadas destinadas al procesamiento gráfico.

Las GPU son una mejora de los chips gráficos monolíticos utilizados en los años 70 y 80. Aunque fue a partir de finales de los años 80 y principios de los 90 cuando se

empezaron a utilizar los microprocesadores de alta velocidad para la implementación de GPU más modernas.

Por un lado, esto permite realizar más tareas mientras se esta ejecutando la red neuronal además de aportar una velocidad inmensamente mayor que ejecutando simplemente con el CPU de cualquier máquina normal. Si se ejecutase la red neuronal solo con la CPU se debería esperar días para que se realizase completamente cualquier acción como podría ser el entrenamiento, cuando con la GPU se obtendria mucha más velocidad permitiendo asi realizar esa misma tarea en apenas varios minutos.

Originariamente se intentó realizar este trabajo desde un ordenador que no disponía de tarjeta gráfica, pero ya desde el comienzo, realizando tutoriales de redes neuronales se podía observar cómo se tardaba horas en ejecutar programas de gran sencillez. Ante esto se vio que era casi imposible realizar este trabajo sin la GPU, ya que a la hora de trabajar con una base de datos grande y una red neuronal compleja los tiempos de espera en las ejecuciones supondrían una espera muy grande.

2.3. CUDA

Computed Unified Device Architecture (Arquitectura Unificada de Dispositivos de Computo) conocido como CUDA [2]es una plataforma de computación en paralelo la cual incluye un compilador y a su vez un conjunto de herramientas de desarrollo. Fue creado el 23 de junio de 2007 por la compañía americana Nvidia y hoy en día están por la versión 9.1. Aunque originariamente está pensado para programación en C gracias a wrappers se pueden usar otros lenguajes de programación como son Fortan, Java y Python.

La función de CUDA es intentar aprovechar al máximo las ventajas de la GPU en contraposición a las de la CPU de propósito general, ya que este ofrece múltiples núcleos, lo cual permiten ejecutar y trabajar con múltiples hilos a la vez realizando cada uno labores independientes (trabajo al que está destinado la GPU) permitiendo así una optimización enorme en el trabajo.

Gracias a esta herramienta ha sido posible este trabajo de fin de grado ya que sin ella no se podría haber optado a utilizar la herramienta de la tarjeta gráfica. Como se ha mencionado en la sección de la GPU, sin ella todo este proyecto habría sido mucho más tedioso y la diferencia de tiempos a la hora de ejecutar habría supuesto una pérdida de días.

2.4. Redes neuronales convolucionales

Las redes neuronales convolucionales[3] son un tipo de red neuronal artificial cuyo aspecto guarda un gran parecido con la corteza visual de un cerebro, donde cada una de las neuronas son los distintos campos receptivos. Este tipo de red neuronal es el más utilizado con respecto a todas las tareas de visión artificial, como en ejemplos como la clasificación y división o segmentación de imágenes que se ha utilizado en este trabajo de fin de grado. Las redes neuronales convolucionales se modelan como colecciones de neuronas que están conectadas entre ellas como un grafo acíclico. Están compuestas por distintas neuronas que tienen pesos y bias aprendibles . Cada una de estas neuronas recibe una entrada, realiza una serie de operaciones y opcionalmente lo conecta de forma no lineal con otra neurona distinta. Los casos más típicos de una arquitectura de CNN es una imagen de entrada que tras aplicarle una serie de capas transforma dicha imagen en una salida sobre la que se tomara una decisión. Hay distintos tipos de capas[4] (layers) véase figura 2.1, de las cuales las más conocidas son:

- Capa convolucional (Convolutional Layer): Realiza una multiplicaci|ón punto a punto entre un volumen y el area seleccionado del nucleo (kernel).
- Capa totalmente conectada (Fully Connected): Las neuronas entre dos capas adyacentes están completamente conectadas entre ellas.
- Pooling Layer: Esta capa introduce un efecto de no linealidad a un sistema que básicamente solo ha estado computando operaciones lineales durante las capas convolucionales.
- Unidad lineal rectificada (ReLu Layer): Se trata de una capa de muestreo descendente que reduce la cantidad de datos. La más típica es la capa de agrupación máxima que mantiene el mayor valor del área seleccionada.

2.5. Deep Learning Frameworks

Un framework[5] o entorno de trabajo en el desarrollo de software es una estructura conceptual y tecnológica de asistencia definida, por lo general, con distintos elementos específicos de software. Los Deep Learning Frameworks tiene su utilidad en que permiten que gente sin unos conocimientos muy avanzados sobre la tecnología de machine learning, puedan descargar y ejecutar redes neuronales, pudiendo así entrenarlas y modificarlas para que realicen un análisis detallado de modelos predictivos que el usuario desee.

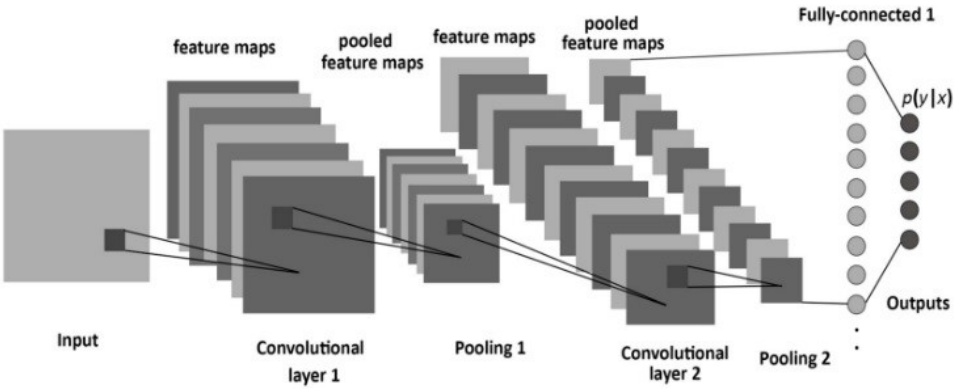


Figura 2.1: Ejemplo de red neuronal convolucional

Los cuatro entornos de trabajo de aprendizaje profundo a día de hoy son:

2.5.1. TensorFlow

Creado por Google [6] , se trata de uno de los entornos de trabajo para Deep Learning más populares a día de hoy, es por ello que es utilizado por grandes compañías como Intel, Twitter o Airbnb. Lleva desde 2015 por lo que desde entonces está triunfando y prosperando de manera activa, recibiendo de forma constante nuevos códigos y actualizaciones.

Se trata del Framework más utilizado por los usuarios actualmente lo cual genera que distintos desarrolladores lo utilicen y generen más documentación y provocando así que sea más sencilla su utilización. Se puede utilizar en distintos lenguajes como C++, Python y R, y permite una gran cantidad de tutoriales explicados de manera muy intuitiva.

2.5.2. Keras

Keras[7] es otro entorno de trabajo que se está expandiendo y está teniendo una rápido adaptación por parte de los usuarios a escala mundial. Una de las principales razones por la que este entorno está triunfando tanto se debe a que tiene una rápida introducción y aprendizaje para usuarios principiantes. Esto se debe a que su arquitectura es muy sencilla de utilizar y por tanto de aprender. Esta implementada para ser utilizada en Python únicamente y puede utilizar redes neuronales convolucionales.

Originariamente Keras se creó como complemento de TensorFlow para poder solucionar los fallos que este entorno generaba en su utilización a los usuarios. Al ser complementario a otro entorno también permite que se pueda ejecutar Keras con

TensorFlow. No necesita de una gran cantidad de código lo cual le coloca como una gran opción para principiantes.

2.5.3. PyTorch

Por otro lado, encontramos PyTorch[8] que se trata de un entorno muchas más nuevo que TensorFlow que está aumentando su popularidad a gran velocidad. Este entorno ha sido creado por Facebook, el cual está aumentando rápidamente su popularidad ya que ha sido implementado para aumentar de forma masiva contando con una gran flexibilidad y pudiendo ser muy personalizable. Es por esto por lo que actualmente este entorno es el segundo más popular a nivel mundial aumentando su popularidad a gran velocidad.

Esta construido sobre Torch, un entorno de computación científica que esta originariamente pensado para el desarrollo de redes neuronales convolucionales. Hace uso de elementos como CUDA u otras librerías de C++ que la hacen más robusta y flexible. Gracias a todos estos elementos PyTorch tiene una arquitectura que permite un desarrollo y entrenamiento de modelos de redes neuronales con gran facilidad.

2.5.4. Caffe

Caffe[9] es un entorno de los mejor valorados ya que se trata de uno de los más rápidos. Esto se debe a que diariamente procesa entorno a unos 60 millones de imágenes únicamente con una GPU K40 de NVIDIA. Otro elemento que provoca que este entorno este muy bien valorado es que funciona con C,C++, Python, Matlab, e incluso CLI. Tiene una arquitectura que permite entrenar redes neuronales sin necesidad de un código excesivamente complejo.

Capítulo 3

Diseño

3.1. Introducción

Al enfrentarnos a este trabajo de fin de grado, pese a que la parte fundamental consistía en la realización de un estudio comparativo de los distintos comportamientos de la red neuronal, el diseño de la misma se mostraba como un punto clave en el desarrollo. Esto es así ya que en función del diseño seleccionado se podría observar diferentes comportamientos, permitiendo la realización de un estudio más o menos profundo sobre la interacción con la red neuronal.

El diseño se dividió en dos aspectos fundamentales. Por un lado la red neuronal a utilizar, contando con las diversas opciones con las que se planteó el trabajo, y sobre la cual se añadirían las capas con las que se realizaría el estudio. Por otro lado, la base de datos con la que se realizaría todo el trabajo, la cual a su vez se dividiría en dos partes, entrenamiento y testeo.

3.2. Diseño red neuronal

Como ya se ha señalado, el diseño de la red neuronal constituía una de las piezas clave de este trabajo, según el cual se cimentarían las bases del resto de estudios en cuestión. Se debe partir de una red neuronal que pudiese ser ejecutada con CUDA, dentro del entorno de PyTorch y que cumpliese el requisito de categorizar lugares. Teniendo esto en cuenta se optó por las siguientes redes:

3.2.1. PLACES365

La red neuronal de PLACES365 [10] diseñada por el MIT y especializada en el reconocimiento de escenas. Esta red toma imágenes o grupos de imágenes y se

entrena para diferenciar todas ellas entre 365 categorías, las cuales van desde 1, airfield(aeropuerto), hasta 365, zen_garden(jardín zen). Para el entrenamiento de esta red neuronal convolucional se ha utilizado la base de datos de PLACES, de la cual se hablará en próximos apartados.

En torno a esta red neuronal se ha desarrollado toda nuestra propia red. Tanto es así que se ha utilizado completamente para la implementación de la misma, a escepción de alguna modificación y añadión de nuevas capas que permitiesen el estudio de aquello que se deseába en este proyecto.

Esta red neuronal originariamente fue implementada para ser utilizada en el entorno de Caffe, pero nuevos entornos válidos han sido añadidos (PyTorch) permitiéndo así el uso de esta red para la realización de este trabajo. Se pudieron encontrar una gran cantidad de redes posibles, algunas más sencillas como AlexNet o ResNet18, y otras más complejas como ResNet50 o DenseNet161. En los anteriores casos el número que los acompaña indica la cantidad de capas que tienen estas redes, siendo ResNet18 una red con 18 capas.

3.2.2. Mi red neuronal

Para el diseño de la red neuronal se ha utilizado la red de Places365. Como ya se ha comentado esta red permite elegir entre 365 categorías, por lo que cumple a la perfección las necesidades para formar la base de nuestra red. También se utilizó el código implementado para realizar el tutorial con el dataset de CIFAR10, que pese a ser este tutorial un sistema de CNN muy simple, ya que solo trabaja con 10 posibles clases y su red es muy básica, también ha servido de base de nuestra implementación.

Posteriormente para diseñar la red deseada se necesitaba reducir esas 365 categorías predeterminadas a tan solo 10 que fueron elegidas en el planteamiento del proyecto. Para ello se añadieron tres capas más a la salida de la red de PLACES365, las cuales van pasando de 365 posibilidades a 90, después a 57 y finalmente a 10.

Otro punto a tener en cuenta en el diseño de nuestra red era el hecho de que debía coger imágenes de nuestra base de datos. Dado que se contába con una GPU, se pudo tomar grupos de imágenesa mayor velocidad. Para ello también se modificó la red anterior que estaba prediseñada para seleccionar archivos comprimidos, pasando a ser ahora grupos de 64 imágenes.

3.3. Diseño de la base de datos

Para el diseño de nuestra red neuronal se tomó como modelo la base de datos de PLACES, pero en un nivel mucho menor. Esta base de datos tiene un papel crucial,

ya que con ella se ha realizado el entrenamiento de la red de PLACES365 utilizada en este trabajo.

3.3.1. Base de datos de PLACES365

El auge de las iniciativas de conjuntos de datos de varios millones de elementos ha permitido que los algoritmos de aprendizaje automático hambrientos de datos alcancen el rendimiento de la clasificación semántica casi humana en tareas como el reconocimiento de objetos visuales y escenas. [11]

Utilizando las redes neuronales convolucionales (CNN) de vanguardia, se proporcionan CNN (lugares-CNN) de clasificación de escenas como líneas de base, que superan significativamente los enfoques anteriores. La visualización de las CNN entrenadas en PLACES muestra que los detectores de objetos emergen como una representación intermedia de la clasificación de escenas. Con su alta cobertura y gran diversidad de ejemplos, la base de datos de lugares junto con las CNN de lugares ofrece un recurso novedoso para guiar el progreso futuro en problemas de reconocimiento de escenas.

3.3.2. Mi base de datos

Al querer entrenar una red neuronal con respecto a un objetivo concreto, se debía disponer de un dataset propio al cual se pudiese acudir para el entrenamiento y posterior testeo de nuestra red neuronal. Sobre este dataset y las imágenes que contiene se ha planteado el estudio de comparación gracias al cual se podrá visualizar el comportamiento real de una red al enfrentarse a imágenes no tan marcadas y con distintas condiciones y características.

Es por ello que en el diseño de nuestra propia base de datos se buscaba que fuesen clases diferenciadas entre sí, pero con puntos en común como pueden ser que todos fuesen de la misma universidad (la Universidad Autónoma de Madrid), o que todos o casi todos fuesen facultades. Otro punto importante era que todas las imágenes fuesen exteriores ya que así podría realizarse un primer estudio donde el ojo humano sí pudiera llegar a distinguir sin necesidad de una capacidad de memoria muy grande, y pudiendo así ver como actuaría nuestra máquina.

Otro punto fundamental del diseño de la red neuronal era dejar marcado de manera muy clara las diferencias entre las imágenes de entrenamiento y de testeo, ya que sin esta diferenciación el posterior estudio de la efectividad de la red estaría completamente adulterado. Además, se buscaba que dentro de la categoría de entrenamiento las distintas categorías también estuviesen diferenciadas para facilitar y permitir el posible entrenamiento.

Capítulo 4

Desarrollo

4.1. Selección del dataset

Como se ha mencionado en el capítulo tres, para la realización de este Trabajo de Fin de Grado se ha hecho uso de dos datasets entorno a los cuales se han realizado todas las pruebas. Para el proyecto que se ha querido realizar, se debía buscar un dataset enfocado hacia el reconocimiento de lugares; por ello, las bases de datos de fotografías de personas o de objetos no eran válidas para este trabajo optándose en su lugar por el uso de los siguientes datasets.

4.1.1. Dataset PLACES

Esta base de datos de imágenes aportada por el MIT (Instituto Tecnológico de Massachussets) es el dataset que se ha empleado para comenzar el desarrollo de este proyecto, cimentando así la creación de nuestra red neuronal.

Esta base de datos de distintos lugares del mundo contiene un repositorio de aproximadamente 10 millones de fotografías y escenas, englobando todo tipo de lugares. Se busca que todas las fotos tengan diferencias con respecto a la perspectiva, momento del día, momento del año en que se toma, a diferentes brillos, con gente de por medio o sin ella, entre otras cosas.

Todo esto hace de este dataset uno de los mejores, por no decir el mejor, para el reconocimiento de escenas estando a libre disposición para cualquiera que desee descargarlo. Pese a que esta base de datos es un pilar fundamental de este proyecto, no ha sido necesario descargarla ni utilizarla. Esto se debe a que, en sí, la red neuronal con la que se ha trabajado, PLACES365, ha sido entrenada originariamente con esta base de datos. No obstante, si fuese necesario entrenar nuestra red desde el principio con ella, se consumiría una enorme cantidad de tiempo, ya que se trata de un dataset

inmenso que contiene en torno a 10 millones de imágenes.

4.2. Dataset propio

Esta es la base de datos fundamental ya que entorno a ella ha girado el grueso de este Trabajo de Fin de Grado. Esto se debe a que este proyecto, como se ha señalado anteriormente, consiste en reconocer escenas de la Universidad Autónoma de Madrid. Para su obtención se partió de la cámara dual de 12 + 5 MP: Bokeh dinámico con IA y autoenfoque dual píxel. Se trata de una cámara de un móvil Xiaomi Redmi Note 6 Pro.

Se tomaron fotos de distintos edificios de la universidad desde distintos ángulos y en diferentes franjas horarias para poder obtener una base de datos eficaz. Este dataset a su vez se divide en dos que son:

TRAIN: Este componente del dataset es el mayor de los dos ya que tiene un total de 250 imágenes diferentes. Para su obtención se realizó fotografías de 10 distintas categorías: rectorado, renfe, plaza mayor, facultades de filosofía, económicas, biología y escuela politécnica superior, el campo de fútbol, el centro nacional de biotecnología y finalmente la residencia de estudiantes. De cada uno de estos edificios se realizaron un total de 25 fotos del exterior, tomando imágenes de distinto tipo, como se ha mencionado anteriormente. Esto proporcionará una mayor variedad a la hora de entrenar la red neuronal mejorando así su preaparación.

TEST: Este es el componente menor del dataset que se ha generado en este proyecto, el cual ha sido utilizado a la hora de testear los resultados de la red neuronal. Para ello se sacaron fotos de los mismos edificios indicados previamente. En este caso se tomaron un total de 10 fotos por edificio, con las mismas características buscadas que para el entrenamiento, es decir, planos exteriores y diferentes unos de otros.

4.3. Selección de la red neuronal

Como se ha mencionado anteriormente este trabajo se basa en la red preentrenada de PLACES365, también creada por el MIT, la cual es originariamente una red neuronal inmensa que toma imágenes de un dataset de interés u otro crado por el MIT. Como resultado se obtiene la probabilidad de que una imagen pertenezca a una de las 365 categorías predefinidas. .

Se ha seleccionado esta red neuronal ya que sirve como base en este trabajo. Partiendo de ella existe la posibilidad de modificar la red para obtener un total de 10 categorías, centradas en edificios de la Universidad Autónoma, frente a las 365 que

hay por defecto.

A su vez ha sido seleccionada porque debido a sus características permitiría realizar un estudio comparativo entre todas las imágenes de nuestro dataset propio, comparándolas con las categorías predeterminadas por la PLACES365. Estos resultados se analizaran mas detalladamente en capítulos posteriores.

4.4. Pruebas con código de PLACES365 sin implementar nuestra red

Primero se realizaron una serie de pruebas en las cuales, partiendo del código inicial que se tenía de PLACES365, se modificó para que, en vez de tomar las imágenes de ejemplo preparadas para el tutorial de esta red neuronal, el código se ejecutase con nuestra base de datos de imágenes de la universidad.

El código de PLACES como ya se ha comentado en apartados previos, está compuesto por una red neuronal que, tras una serie de operaciones computacionales, termina dándote las probabilidades de que cada imagen que pase por la red sea una de las 365 categorías que tiene predefinidas este código. En este caso se quería modificar la implementación predefinida, para que, sin alterar la red neuronal fuese posible cambiar el dataset del que obtenía las imágenes, y con el que se realizaba la parte de testeo.

A continuación, dichas probabilidades se fueron guardando en archivos de texto, para realizar un estudio comparativo de las distintas probabilidades en Matlab. De cada categoría del dataset se guardaban por separado los resultados obtenido tras el testeo con la red neuronal.

Finalmente se realizó un estudio mediante Matlab de las distintos carpetas de la base de datos con la finalidad de poder observar si cada categoría guardaba relación entre todas sus imágenes, y también si las distintas categorías guardaban algún tipo de similitud entre ellas. Todos los resultados obtenidos en este apartado están comentados en el próximo capítulo.

4.5. Entrenamiento de la red neuronal

El entrenamiento de la red neuronal ha sido uno de los aspectos fundamentales en el desarrollo de este proyecto. Para ello como ya se ha comentado en el capitulo anterior, se ha utilizado en primer lugar la red ResNet18 de PLACES365. Se ha hecho transfer learning añadiendole tres capas nuevas al modelo existente. Las capas se situaron después del modelo ya predefinido, donde partiendo de las 365 categorías

que te daba como salida esta red, se reducía en la primera capa a 120, en la siguiente capa hasta 84 y finalmente a 10 categorías, siendo estas últimas los edificios de las distintas facultades de la universidad.

Para la realización del entrenamiento se modifico el dataset del que se obtenian las imágenes cambiando el de PLACES, por el dataset propio comentado en el capítulo anterior. De este dataset se seleccionarían las escenas en grupos de 64 trabajando con 4 hilos a la vez, logrando así una mayor velocidad de ejecucción. Lograr esta velocidad de ejecucción se debe a que contabamos con la herramienta de CUDA descrita en el capítulo dos.

Por otro lado, también se probo a realizar el entrenamiento de nuestra red utilizando como base en esta ocasión ResNet50 de PLACES365. Esto se realizó para conseguir un entrenamiento con una red mas completa que la de ResNet18, consiguiendo de esta manera unos resultados más efectivos.

Todos los resultados obtenidos, tanto para las pruebas con ResNet18 como para los resultados obtenidos con ResNet50 estan comentados en el siguiente capítulo.

4.6. Testeo de la red neuronal

Para la realización del testeo de la red neuronal primero se han obtenido los pesos guardados previamente en el entrenamiento .

Una vez se tienen los pesos guardados se cambia la selección del dataset, para que en este caso seleccione el de test y también se modifica de manera que trabaje con todas las imágenes a la vez en un único hilo. La red se ejecutará de nuevo, con los pesos ya guardados y realizará el testeo. Se ha desarrollado de manera que pase por pantalla el porcentaje de acierto total, el porcentaje de acierto de cada una de las categorías por separado y finalmente el estudio imagen a imagen. El tiempo de ejecución de la red neuronal es bastante breve, siendo en torno a un minuto de duración la operación de testeo completa. Los resultados obtenidos en este apartado se analizarán en el siguiente capítulo.

Capítulo 5

Integración pruebas y resultados

5.1. Primeras pruebas

Las primeras pruebas como ya se ha comentado anteriormente, se han realizado sobre la red ya preentrenada Places365. Para ello se ha alterado la base de datos de la que se obtenían las imágenes para realizar el testeo, especificando que el dataset fuese el nuestro en vez del predeterminado.

Esta primera prueba se ha realizado para observar cómo reacciona la red, no ante imágenes de archivo, sino con imágenes de edificios de un campus universitario . Ante este dato, sabemos que la mayor parte de escenas deberían dar la categoría 78, es decir la categoría campus predeterminada de la red preentrenada Places365, o bien categorías como la 338, que es, la de la estación de tren, para el caso, la estación de Renfe .

Primero las pruebas se realizaron con las distintas imágenes de cada categoría dentro de la selección de entrenamiento, es decir, primero las 25 escenas de la categoría 1, que en este caso era la Escuela Politécnica Superior, posteriormente la categoría 2, el Centro Nacional de Biotecnología, y así sucesivamente hasta trabajar con todas las categorías. Una vez obtenidas las probabilidades se ha podido observar, gracias a distintos elementos de Matlab, como en su mayoría estas imágenes no tenían los mismos resultados, pese a ser los mismos edificios.

Para el caso de la Escuela Politécnica Superior se han obtenido varios picos en distintas categorías (Figura 5.1), siendo los tres más destacados: las categorías 179 con una probabilidad de 0.497, y las categorías 78 y 50, ambas con una probabilidad de 0.434.

Gracias a esta gráfica he podido observar como, por un lado, las probabilidades no son altas, ya que ninguna supera el cincuenta por ciento de compatibilidad

con ninguna de las clases; y por otro lado, que ofrece una variedad muy grande de categorías.

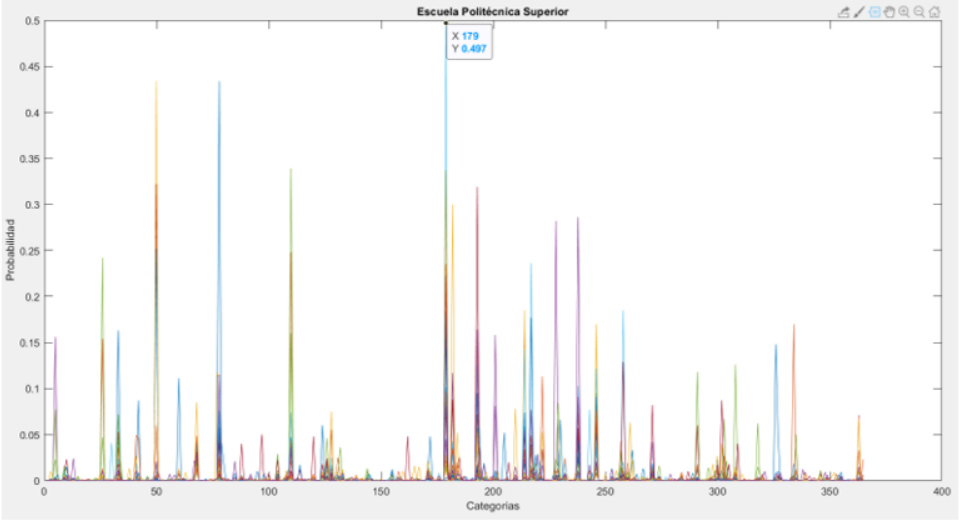
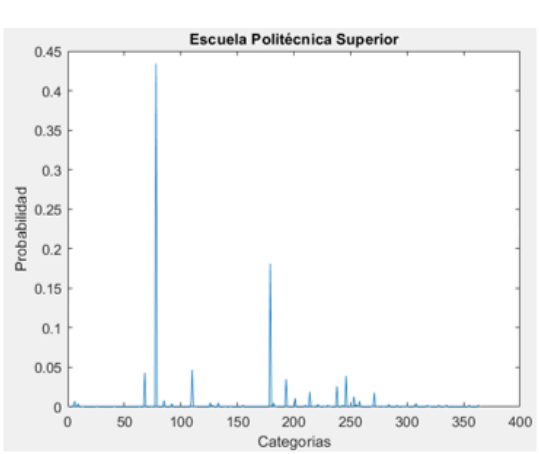


Figura 5.1: Gráfica de la carpeta completa de EPS

Se ha observado que la mayor probabilidad de compatibilidad coincide con la categoría 179, que se corresponde con la categoría hospital. También se ha visto que las dos siguientes categorías que más similitud comparten, son la categoría 50, que corresponde con casa de playa (beach house), y la 78 que corresponde con campus.



(a) Gráfica categoría 78 EPS

(b) Imagen EPS

Figura 5.2: Gráfica categoría 78 EPS 5.2(a) e Imagen EPS 5.2(b)

Ante estos primeros resultados se ha podido deducir que el nivel de acierto no es alto, ya que la probabilidad de similitud con la categoría correcta (campus) es en

este caso de 0.434, inferior a un cincuenta por ciento de coincidencia. Además, se ha podido observar que destacan más otras categorías como la de hospital, reflejando el error a la hora de categorizar.

Por otro lado, se ha realizado el estudio para las imágenes del campo de futbol de la universidad, obteniendo la siguiente gráfica (Figura 5.3):

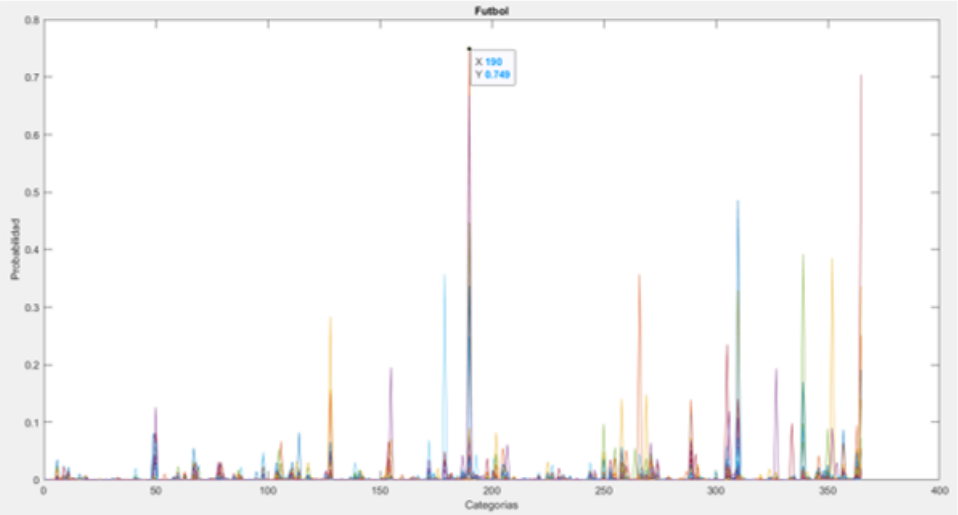


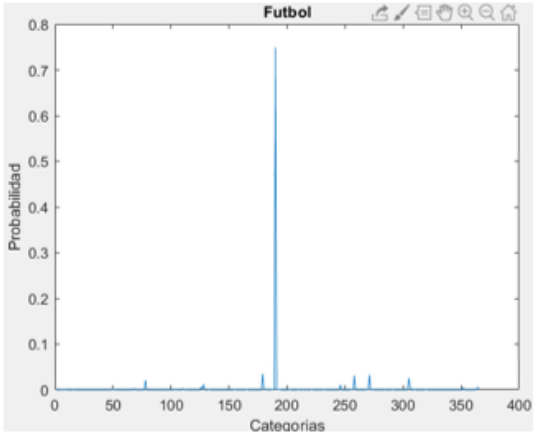
Figura 5.3: Gráfica de la carpeta completa fútbol

Como se ha podido observar se tienen dos categorías que destacan por encima del resto, que son la 190, con una probabilidad de 0.749, y la categoría 365, con una probabilidad de 0.704. Se han descartado el resto de las categorías, al estar todas por debajo del 0.5 de compatibilidad. De esto se ha podido deducir que en esta grafica nos encontramos muchas imágenes que aportan información errónea, pero que sin embargo es posible encontrar dos picos destacables.

La categoría con mayor compatibilidad es la 190, que se corresponde con pista exterior de patinaje sobre hielo (ice_skating_rink/outdoor) (Figura 5.4). He supuesto que este fallo se debe a, o bien que al utilizar fotos como las siguientes, donde el campo de futbol es de arena y con un nivel de luz muy alto, la red supone que se trata de un campo de patinaje sobre hielo; o bien lo confunde con una pista de patinaje seca, donde en lugar de hielo encontraríamos arena.

Por otro lado, la otra categoría que más destaca es la 365, que se corresponde con jardín zen, se ha analizado en que imagen se producía el fallo y era en la siguiente (Figura 5.5):

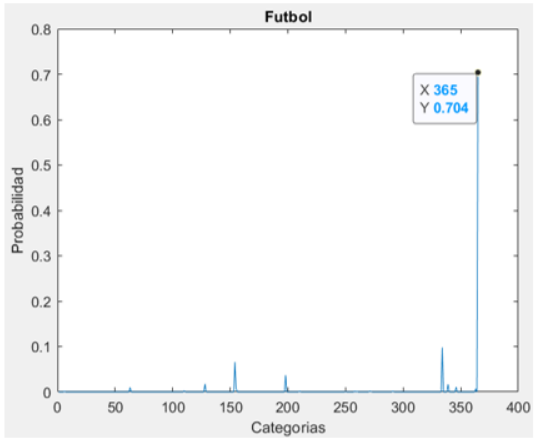
Para este error no se ha logrado alcanzar una idea clara del motivo por el que confunde la imagen, aunque se ha supuesto que la causa



(a) Gráfica categoría 190 fútbol

(b) Imagen 1 campo de fútbol

Figura 5.4: Gráfica categoría 190 fútbol 5.4(a) e imagen 1 campo de fútbol 5.4(b)



(a) Gráfica categoría 365 fútbol

(b) Imagen 2 campo de fútbol

Figura 5.5: Gráfica categoría 365 fútbol 5.5(a) e imagen 2 campo de fútbol 5.5(b)

principal del fallo es debida a la gran cantidad de vegetación que vemos en la foto, mezclada con arena.

Después se ha comprobado que el mayor pico de todos se ha encontrado cuando testeamos con la carpeta de Plaza Mayor. Al realizar el testeo con todas las imágenes de esta carpeta se ha obtenido la siguiente gráfica (Figura 5.6):

En esta grafica se observa como claramente hay un pico de similitud que destaca sobre el resto, el indicado en la imagen, seguido de un pico menor pero también destacable, que se corresponde con la categoría 3 y otro pico más en la categoría 189. El resto de picos se han descartado al ser mucho menores que los ya mencionados.

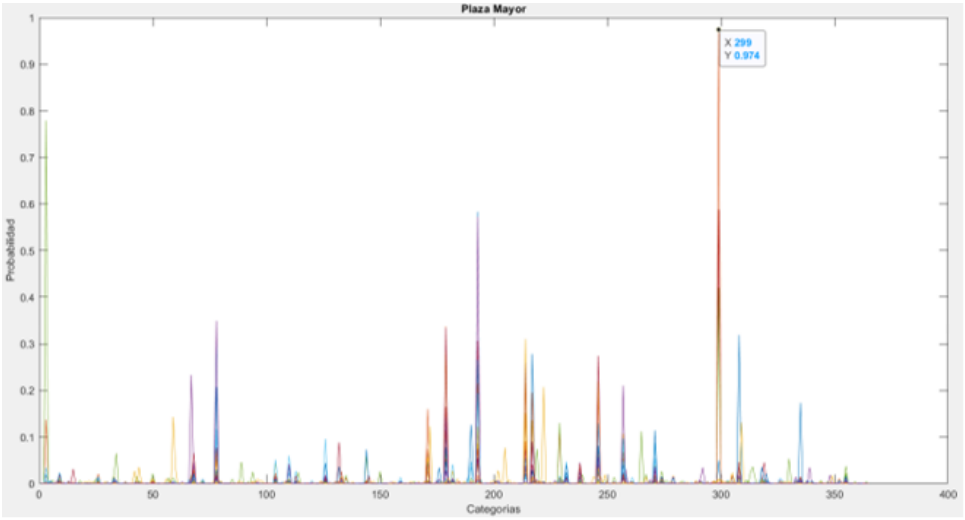
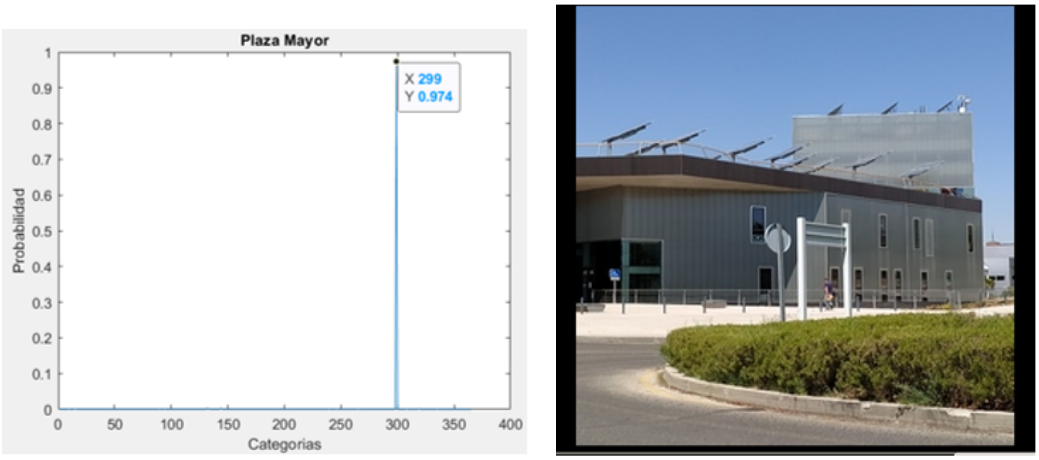


Figura 5.6: Gráfica de la carpeta completa Plaza Mayor

Se ha aislado la categoría que destaca sobre el resto con una probabilidad de 0.974, que como se ha dicho es la 299, que corresponde a una habitación llena de servidores (server room). Por otro lado, la categoría 3, con una probabilidad de 0.779, corresponde con una terminal de aeropuerto (Figura 5.7).



(a) Gráfica categoría 299 Plaza Mayor (b) imagen de la Plaza Mayor

Figura 5.7: Gráfica categoría 299 Plaza Mayor 5.7(a) e imagen de la Plaza Mayor 5.7(b)

De esto se ha deducido que la red falla estrepitosamente al analizar las imágenes de esta categoría, ya que ambas categorías son erróneas y no encuentra en ningún momento similitud con lo que es en realidad, que seria, centro comercial. En general se

ha podido observar que, al realizar el estudio de esta red neuronal sin un entrenamiento previo con imágenes del campus de la universidad, el nivel de fallo es altísimo, dentro de que como se ha podido observar con la gráfica total (Figura 5.8, se han obtenido probabilidades muy dispersas unas de otras y de distintos elementos.

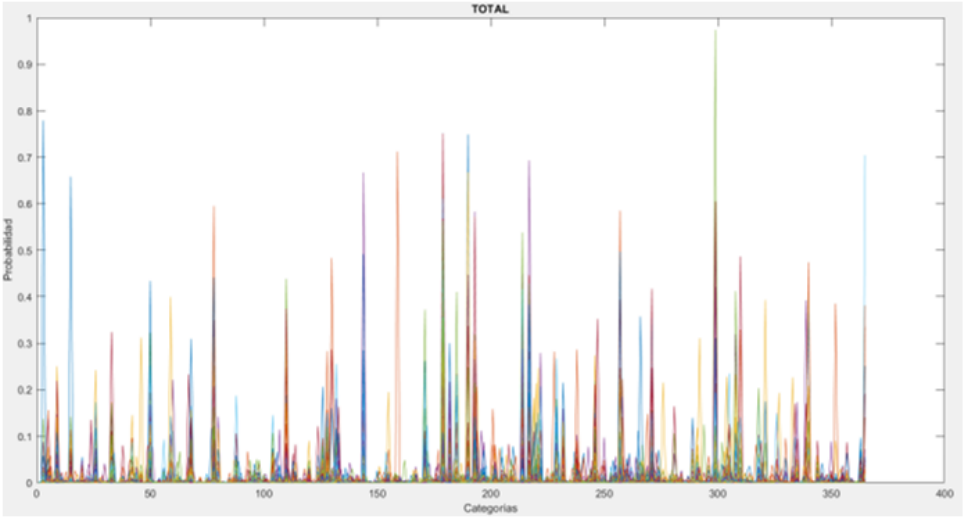


Figura 5.8: Gráfica de todas las imágenes

5.2. Resultados del entrenamiento

Como resultado del entrenamiento (Figura 5.9) podemos observar que la red neuronal tras 1000 iteraciones y un entrenamiento completo con nuestra propia base de datos, nos da un error bastante bajo de 0.00002. Podemos observar cómo al comienzo del entrenamiento el error era altísimo, en torno a 2.26119 y tras un tiempo de ejecución alrededor de unos 20 minutos logra bajarlo a un éxito casi completo.

Se puede observar como la red neuronal reduce en gran medida el nivel de error en las primeras 20 iteraciones pasando desde el máximo ya comentado a casi 0.04471. Esto supone que a partir de 20 iteraciones contando con que selecciona grupos de 64 imágenes ya ha podido trabajar multitud de veces con las imágenes del dataset, ya que nuestra base de datos en total cuenta con 250 imágenes, y la red neuronal ha interactuado ya con 1280 imágenes aproximadamente.

Un detalle que se pudo observar es que a partir de la iteración 800 aproximadamente el error iba variando únicamente entre 0.00002 y 0.00001, por lo que suponemos que ya a partir de este punto la red ha alcanzado un máximo en su entrenamiento y simplemente varía el resultado en función del grupo de imágenes que este seleccionando. En ambos casos la probabilidad de error es tan baja que se podría considerar

casi nula y la red estaría cumpliendo los objetivos deseados.

Este resultado nos induce a pensar que nuestra red neuronal debería estar preparada para enfrentarse a cualquier imagen de las distintas facultades y edificios con los que se ha trabajado, es decir que no debería tener error ante ninguna imagen de prueba.

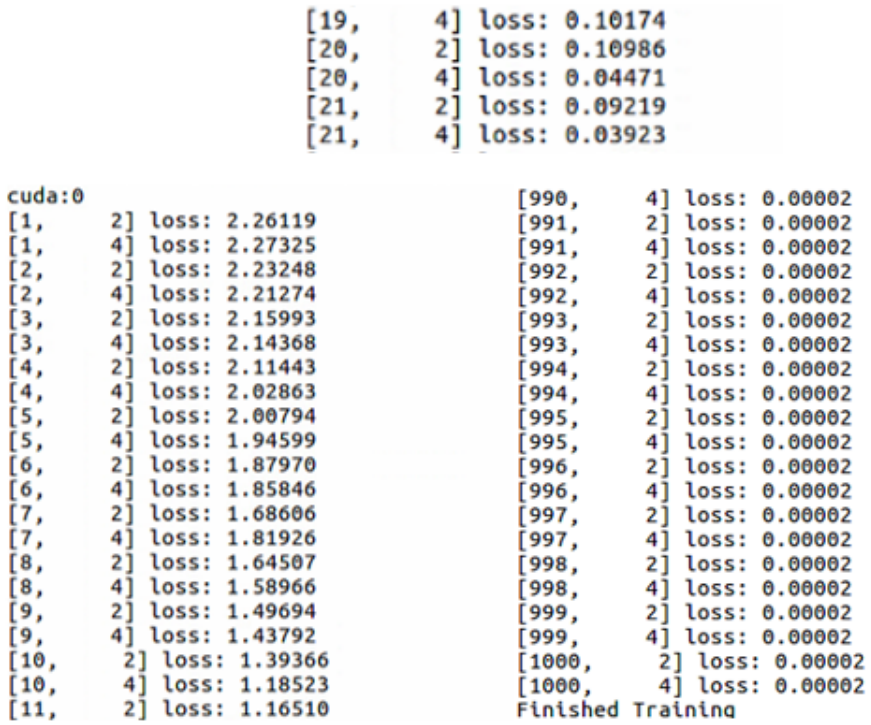


Figura 5.9: Resultados por pantalla del train

5.3. Resultado del testeo

Tras la realización del testeo se obtuvieron los siguientes resultados (Figura 5.10):

Figura 5.10: Porcentaje total por pantalla del test

Lo primero que se puede observar de estos resultados es el porcentaje obtenido

de acierto total con las 1000 imágenes de test, siendo igual a un 82 %. Se estiman los siguientes intervalos para los valores de test, todos ellos en porcentaje:

- [50,60): Resultado obtenido es malo.
- [60,75): Resultado obtenido es medio.
- [75,90): Resultado obtenido es bueno.
- [90,97): Resultado obtenido es muy bueno.
- [97,100): Resultado obtenido es excelente.

Sabiendo esto se puede estimar que el resultado obtenido en este testeo de la red neuronal es bueno. Esto nos hace suponer que tiene un rango de mejora bastante alto, ya que se encuentra en una posición bastante media en la tabla de resultados, pero que al tratarse de la primera vez que se realiza el testeo con esta red los resultados son bastante buenos.

```
Accuracy of   EPS : 90 %
Accuracy of BIOTEC : 80 %
Accuracy of RENFE : 60 %
Accuracy of RECTOR : 80 %
Accuracy of PMAYOR : 90 %
Accuracy of DERECHO : 80 %
Accuracy of FUTBOL : 100 %
Accuracy of   ECO : 80 %
Accuracy of  FILO : 70 %
Accuracy of  RESI : 90 %
```

Figura 5.11: Porcentaje de aciertos por separado del test

Por otro lado, también se ha realizado un estudio sobre el porcentaje de acierto de cada categoría en concreto dentro de las distintas carpetas del dataset de testeo (Figura 5.11). La única carpeta con un éxito completo es la carpeta fútbol, con un 100 % de acierto; seguida de las carpetas Escuela Politécnica Superior (EPS), Residencia (RESI) y Plaza Mayor (PMAYOR), con un porcentaje muy bueno de acierto del 90 %. Dentro de la media en la que se encuentra la red en general también encontramos las carpetas del Centro Nacional de Biotecnología (BIOTEC), Rectorado (RECTOR), y las facultades de derecho (DERECHO) y economía (ECO), todas con un porcentaje de acierto del 80 %. Finalmente, las dos carpetas que se pueden considerar como un fallo son la facultad de filosofía (FILO), con un 70 %, y la estación de Renfe (RENFE) con un 60 %.

A continuación, se ha realizado un estudio más detallado de cada una de las carpetas, tratando de ver en que fotos se encuentran los fallos. Por un lado tendremos *GroundTruth*, que representará cada una de las imágenes dentro de la carpeta; y por el otro lado, tendremos *Predicted*, que mostrará la predicción obtenida por la red.

5.3.1. EPS

```
GroundTruth:    EPS   EPS   EPS   EPS   EPS   EPS   EPS   EPS   EPS   EPS
Predicted:    EPS   EPS   EPS   EPS DERECHO   EPS   EPS   EPS   EPS   EPS

GroundTruth:  BIOTEC BIOTEC BIOTEC BIOTEC BIOTEC BIOTEC BIOTEC BIOTEC BIOTEC BIOTEC
Predicted:  RECTOR BIOTEC BIOTEC BIOTEC  FILO BIOTEC BIOTEC BIOTEC BIOTEC BIOTEC

GroundTruth:  RENFE RENFE RENFE RENFE RENFE RENFE RENFE RENFE RENFE RENFE
Predicted:  RENFE  FILO RENFE RENFE RENFE RENFE RENFE  FILO BIOTEC  FILO

GroundTruth:  RECTOR RECTOR RECTOR RECTOR RECTOR RECTOR RECTOR RECTOR RECTOR RECTOR
Predicted:  RECTOR RECTOR RECTOR DERECHO RECTOR RECTOR RECTOR RECTOR RECTOR  FILO

GroundTruth:  PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR
Predicted:  PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR PMAYOR  RESI PMAYOR PMAYOR

GroundTruth:  DERECHO DERECHO DERECHO DERECHO DERECHO DERECHO DERECHO DERECHO DERECHO DERECHO
Predicted:  DERECHO DERECHO DERECHO DERECHO   EPS DERECHO DERECHO DERECHO DERECHO   ECO

GroundTruth:  FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL
Predicted:  FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL FUTBOL

GroundTruth:    ECO   ECO   ECO   ECO   ECO   ECO   ECO   ECO   ECO   ECO
Predicted:  DERECHO   ECO   ECO  FILO   ECO   ECO   ECO   ECO   ECO   ECO

GroundTruth:   FILO  FILO  FILO  FILO  FILO  FILO  FILO  FILO  FILO  FILO
Predicted:   FILO RECTOR  FILO  FILO  FILO  FILO  FILO  RESI  RESI  FILO

GroundTruth:   RESI  RESI  RESI  RESI  RESI  RESI  RESI  RESI  RESI  RESI
Predicted:   RESI  RESI  RESI  RESI  RESI BIOTEC  RESI  RESI  RESI  RESI
```

Figura 5.12: Testeo de cada imagen y su resultado

Como se puede observar esta carpeta solo encuentra un fallo en la quinta imagen mostrada en la Figura 5.13.

Figura 5.13: Imagen de test EPS

Se ha supuesto que este error se debe a que tanto la EPS como la facultad de Derecho están construidas con ladrillo naranja, pudiendo así provocar un error ente ambos edificios. A continuación, también se muestra una imagen de la facultad de derecho 5.14 para visualizar con mayor claridad el error.

Figura 5.14: Imagen de train de derecho

5.3.2. Biotecnología

En este caso concreto se puede ver en la figura 5.12 que el error se encuentra en la primera y la quinta imágenes, donde las confunde con el edificio de rectorado y la facultad de Filosofía.

Como se puede ver en la siguiente figura 5.15 es fácil llegar a entender el motivo por el que nuestra red confunde ambos edificios.

(a) Centro Nacional de Biotecnología test (b) Rectorado train

Figura 5.15: Comparación del fallo en la carpeta de Biotecnología

5.3.3. Renfe

Este caso se podría tratar como un fallo dentro de la red, donde el nivel de acierto es de tan solo un 60 % como se observa en la figura 5.11. Para entender mejor este caso se ha comprobado en qué imágenes ha errado, viendo que en su mayoría se confunden entre Renfe y la facultad de Filosofía (ver 5.12).

A continuación, se muestra la imagen de test de la categoría Renfe y la imagen de entrenamiento (Figura 5.16), con las cuales se puede suponer que la red genera el error correspondiente a la facultad de Filosofía.

5.3.4. Rectorado

En este ejemplo hemos encontrado un fallo del 20 %, concretamente en las imágenes 4 y 10, véase 5.12. Para el caso del error con derecho se ha supuesto que se deba a la similitud entre el ladrillo marrón de rectorado y el ladrillo naranja con un nivel

(a) Imagen de test de Renfe (b) Imagen de train Facultad de Filosofía

Figura 5.16: Comparación del fallo en carpeta de Renfe

de brillo bajo(Figura 5.17). Para el otro ejemplo en cuestión se ha supuesto que se debe a la vegetación que pueda aparecer en la foto.

(a) Imagen de test de Rectorado (b) Imagen de train facultad de Derecho

Figura 5.17: Comparación del fallo en carpeta de Rectorado

5.3.5. Plaza Mayor

Este ejemplo ha tenido un éxito casi completo a excepción de la imagen 8 (ver 5.12), donde se suponía que se trataba de la residencia en vez de la categoría correcta que debería ser Plaza Mayor.

Como se muestra en la figura 5.18 el error lo asume al tratarse ambos edificios de un contrachapado metálico, que reflejado con la luz solar tiene un aspecto de color plateado.

(a) Imagen de test de Plaza Mayor

(b) Imagen de train de residencia

Figura 5.18: Comparación del fallo en carpeta de Renfe

5.3.6. Derecho

En este ejemplo la confusión es también mínima, de solo una imagen. El error se produce con una escena de la Escuela Politécnica Superior que como ya se ha comentado en otro apartado de este mismo capítulo, tiene un aspecto muy similar a esta facultad véase figuras 5.13 y 5.14 .

5.3.7. Fútbol

En esta carpeta se ha obtenido un 100 % de acierto. A continuación, se muestra un ejemplo (figura 5.19) de esta carpeta donde se puede ver claramente como al tratarse de un tipo de escena completamente distinto al resto de categorías, no se encuentra ningún tipo de error. Esta imágenes es un terreno de arena con árboles, pudiendo ser descritas como escenas naturales, frente al resto de escenas que muestran imágenes urbanas.

5.3.8. Economía

Este ejemplo es interesante porque confunde dos escenas con dos facultades bastante distintas entre ellas (ver figura 5.12), ya que la de derecho es una facultad de

Figura 5.19: Imagen de train de campo de fútbol

ladrillo naranja, mientras que la de filosofía es una facultad de color azul o gris, y como se muestra en la imagen 5.20(a) la facultad de económicas es de chapa naranja

Por un lado, se ha supuesto que la confusión con la facultad de derecho se debe a que ambas facultades tienen un tono naranja. Por otro lado, la confusión con la

(a) Imagen de test de económicas

(b) Imagen de train de derecho

Figura 5.20: Comparación del fallo en carpeta de Economía

facultad de filosofía se cree que se debe a la imagen 5.21 de la carpeta de entrenamiento de dicha facultad. Se puede observar como en esta facultad encontramos una puerta y una ventana de color naranja que han generado este error.

Figura 5.21: Imagen de train de filosofía

5.3.9. Filosofía

En esta parte del test hemos tenido un fallo considerable, de hasta tres imágenes, donde dos de ellas las confunde con la residencia de estudiantes y la otra imagen con el rectorado.

Como podemos observar en la figura 5.22, al ser ambas facultades de tonos apagados, tirando a grises, y al ser imágenes con un brillo alto se puede llegar a entender el motivo del fallo.

(a) Imagen 1 de test de filosofía

(b) Imagen de train de residencia

Figura 5.22: Comparación del fallo en carpeta de Filosofía

En cambio, el error con el edificio de rectorado no se llega a entender muy bien ya que como se muestra en la imagen 5.23 del testeo no hay similitudes posibles, salvo quizás por las plantas.

Figura 5.23: Imagen 2 de test de filosofía

5.3.10. Residencia

En este último caso solo hemos encontrado una imagen con error figura 5.12.

Se ha supuesto que este error se debe a que ciertas imágenes del Centro Nacional de Biotecnología al tener un nivel de brillo muy alto las franjas verdes, pueden parecer más oscuras, pudiendo así ser confundida con la residencia.

(a) Imagen de test de residencia (b) Imagen de train de biotecnología

Figura 5.24: Comparación del fallo en carpeta de Residencia

Capítulo 6

Conclusiones y trabajo futuro

6.1. Conclusiones

En el presente trabajo se ha realizado un estudio comparativo de los distintos resultados que se pueden obtener al trabajar con redes neuronales. En un caso sin aplicarle un entrenamiento previo y sin especificar las categorÃas deseadas, y en el otro caso haciendo una selección de las categorías y realizando un entrenamiento.

Se ha logrado analizar los distintos comportamientos que tiene la red, mostrando mediante ejemplos claros cómo la efectividad obtenida con una red preparada para un ejemplo concreto ,como el de PLACES365, (en el cual se realiza un entrenamiento previo con un dataset ajeno al utilizado en pruebas posteriores) tiene una gran cantidad de fallos. Esto evidencia el hecho de que una red preparada para un caso de reconocimiento de escenas no tiene porqué estar preparado para acertar en casos tan concretos como los propuestos en este mismo proyecto. No obstante, con un dataset pequeño propio, pero siendo entrenado con una red modificada en base a una salida deseada, se obtienen unos buenos resultados en un primer testeo, alcanzando una tasa de acierto total de un 82 %.

Asimismo, este Trabajo de Fin de Grado ha sido de gran utilidad para aprender y poner en práctica lenguajes de programación los cuales no había tenido la posibilidad de tratar aún, como es el caso de Python y más concretamente la librería de PyTorch. De la misma manera, me ha permitido avanzar en mis conocimientos sobre las redes neuronales y poder enfrentarme de esta manera a un ejemplo practico de trabajo de inteligencia artificial.

En conclusión, este trabajo ha fomentado el aprendizaje de un lenguaje y los distintos elementos para trabajar con una red neuronal, junto con un mayor conocimiento sobre estudios comparativos y las redes neuronales.

6.2. Trabajo futuro

A continuación se pasará a enunciar posibles pruebas futuras o posibles mejoras con las que perfeccionar elementos de este proyecto:

- Mejorar la sección de testeo de la red neuronal o bien mejorar las carpetas de donde obtiene la escenas e incluso tratar de buscar modificar algunas capas de la red neuronal.
- Realizar un estudio comparativo de los resultados obtenidos no solo con las redes ResNet18 y ResNet50, sino también añadir redes como AlexNet o DenseNet161.
- Aumentar el dataset de escenas de la universidad, creando nuevas carpetas con más edificios y añadiendo distintas imágenes de los edificios ya existentes.
- Una vez aumentada la base de datos tratar de añadir escenas interiores de los distintos edificios y ver qué resultados se obtienen tras un entrenamiento previo.
- Realizar este mismo trabajo con otros frameworks como Keras o TensorFlow, y realizar un estudio para comprobar distintos resultados.
- Tratar de lanzar una aplicación para estudiantes de la Universidad Autónoma basada en este proyecto que les sirva de guía dentro del campus.

Bibliografía

[1] I. S. Ufimtsev and T. J. MartÃnez, "Graphical processing units for quantum chemistry," *Computing in Science Engineering*, vol. 10, pp. 26–34, Nov 2008.

[2] M. Al-Mouhamed, A. Khan, and N. Mohammad, "A review of cuda optimization techniques and tools for structured grid computing," *Computing*, 07 2019.

[3] S. Bosse, D. Maniry, K.Œller@, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, pp. 206–219, Jan 2018.

[4] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, pp. 4–27, March 1990.

[5] S. Tokui and K. Oono, "Chainer : a next-generation open source framework for deep learning," 2015.

[6] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[7] F. Chollet *et al.*, "Keras." https://keras.io, 2015.

[8] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[9] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[10] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[11] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 487–495, Curran Associates, Inc., 2014.